

DT #46526 QA:NA 02/03/2006

MOL.20060309.0023



OCRWM Office of Science and Technology and International



# ***Data Mining Of Experimental Corrosion Data Using Neural Network***

**208th Electrochemical Society Meeting**

**M. KamrunNahar and Mirna Urquidi-Macdonald  
Penn State University**

October 17, 2005  
Los Angeles, CA

This presentation has been funded in whole or in part by the U.S. Department of Energy

# Acknowledgment

## Data Mining Of Experimental Corrosion Data Using Neural Network

M. KamrunNahar and Mirna Urquidi-Macdonald

Penn State University

Support by the Office of Science and Technology and International of the U.S. Department of Energy, Office of Civilian Radioactive Waste Management is gratefully acknowledged. The work was performed under the Corrosion and Materials Performance Cooperative, DOE Cooperative Agreement Number: DE-FC28-04RW12252

The views, opinions, findings, and conclusions or recommendations of authors expressed herein do not necessarily state or reflect those of the DOE/OCRWM/OST&I.



# Research Objectives

- > Establish a data mining methodology for predicting metal/alloy life over extended periods of time
- > Demonstrate the predictive capabilities of data mining by using data available for alloy materials to predict their performance and compare our predictions to the prediction of deterministic models
- > Establish a methodology on data representation and analysis that can be used as a complementary tool in the laboratory to avoid time consuming capturing of additional data
- > Learn and generalize material behaviors based on existing data

# Data Mining

Data mining (also known as Knowledge Discovery in Databases - KDD):

- The nontrivial extraction of implicitly useful information from data
- It uses machine learning, statistical, visualization, and neural network (NN) techniques to discover and present knowledge embedded in data in a form which is easily comprehensible to humans



# Activities in Data Mining

- **Data collection**
  - > (i) Data collection and formatting
  - > (ii) Data cleaning
  - > (iii) Data integration
  - > (iv) Data transformation
- **Data analysis and model development**
  - > (v) Data mining or knowledge extraction
  - > (vi) Pattern evaluation
- **Knowledge presentation**

# Types of Data Analysis

- **Directed data mining**

- Classification
- Estimation
- Prediction

- **Undirected data mining**

- Affinity grouping or association rules
- Clustering

**Our focus:** Prediction data mining

**System of interest:** Corrosion behavior of metals/ alloys

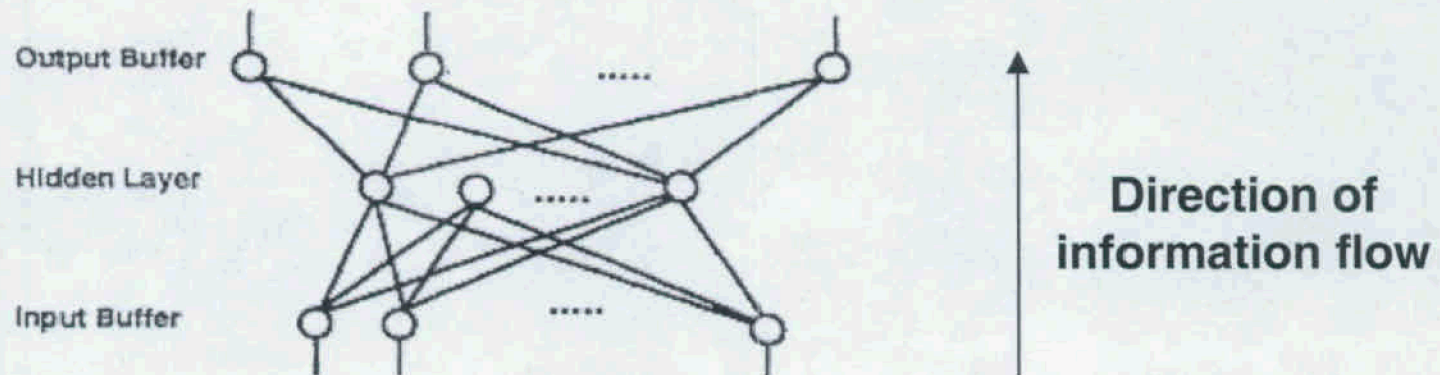
**Method of analysis:** Neural network backpropagation



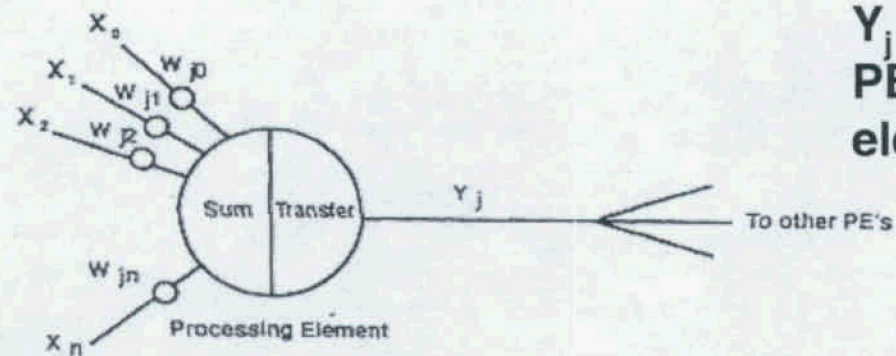
# Neural Network

- **Analogous to human brain activity**
  - > We have neurons in our brain
  - > We learn through the neuron interconnections (stronger or weaker)
- **Tasks that the brain performs well**
  - > Pattern recognition
  - > Classification
  - > Categorization
  - > Pattern reconstruction
  - > Association
- **Neural network model**
  - > No model is imposed
  - > It learns from data

# Neural Network



Typical neural network (NN)



$x_i$  -- Inputs  
 $w_i$  -- Weights  
 $Y_j$  -- Outputs  
 PE -- Processing element

Processing element

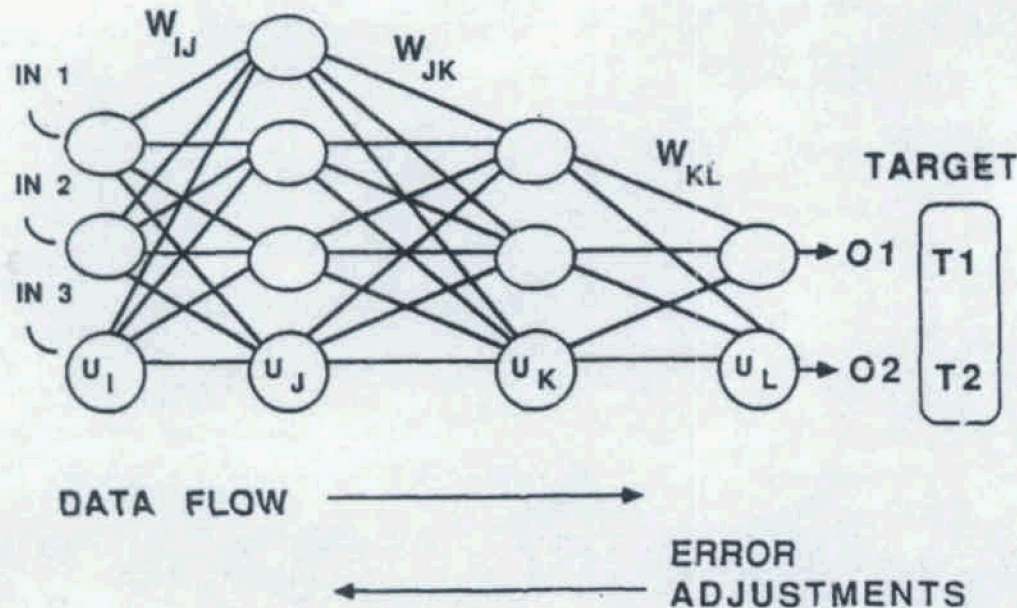


# BackPropagation Neural Network

- BackPropagation (BP) Neural Network (NN) is a learning algorithm that learns by example where the user provides a set of inputs and known outputs
- The BP learning process works in iterative steps:
  - > The network initially starts with one example of inputs and some random weights and produces outputs based on the current state of the inputs and weights
  - > The outputs are compared to the known outputs
  - > The error (mean-squared) between the network outputs and the known outputs is propagated backwards through the network and small changes are made to the weights in each layer to reduce the error for the case in questions
  - > The whole process is repeated for each of the example cases and back to the first case again and so on
  - > The cycle is repeated until the overall error value drops below some pre-specified error threshold
  - > The network is then said to have learned the mapping functions between the inputs and outputs and the final weights are calculated

# Neural Network

To extrapolate and establish relationships to categorize importance of input parameters (pH, T, [Cl<sup>-</sup>], etc.) on output parameters (general corrosion rate, localized damage, etc.), neurons and neuron layers can be simulated mathematically: supervised learning with backpropagation.



Typical backpropagation layout



# Our Approach

- > Data collection on generalized and localized corrosion of several metals in contact with several environments
- > Data reprocessing, that includes data filtering (for bias information), data transformation (to accentuate certain data features) and data normalization (to give all variables the same order of importance)
- > Statistical data “cleaning” used to identify outliers
- > Data representation in different spaces using mathematical transformations; it is a very important step to identify contrasting data features
- > Data analysis
  - >> Identify hidden functions that relate measured parameters to alloy corrosion rates. The discovered function (supervised backpropagation in neural network) is used to make extrapolations

# Data Mining Tasks

- Data collection

- > General corrosion data on Iron-alloy ( $\text{Fe}_{68}\text{Ni}_{14-x}\text{Mo}_x\text{Si}_2\text{B}_{16}$  metallic glasses for different x values) (source: article available publicly [1])
- > Crevice corrosion on Ti-alloy (grade-2 Titanium) (source: article available publicly [2])

- Data processing and model development

- > Backpropagation designed for predicting the polarization behavior for metallic glasses
- > Backpropagation developed for predicting localized corrosion behavior for grade-2 Titanium

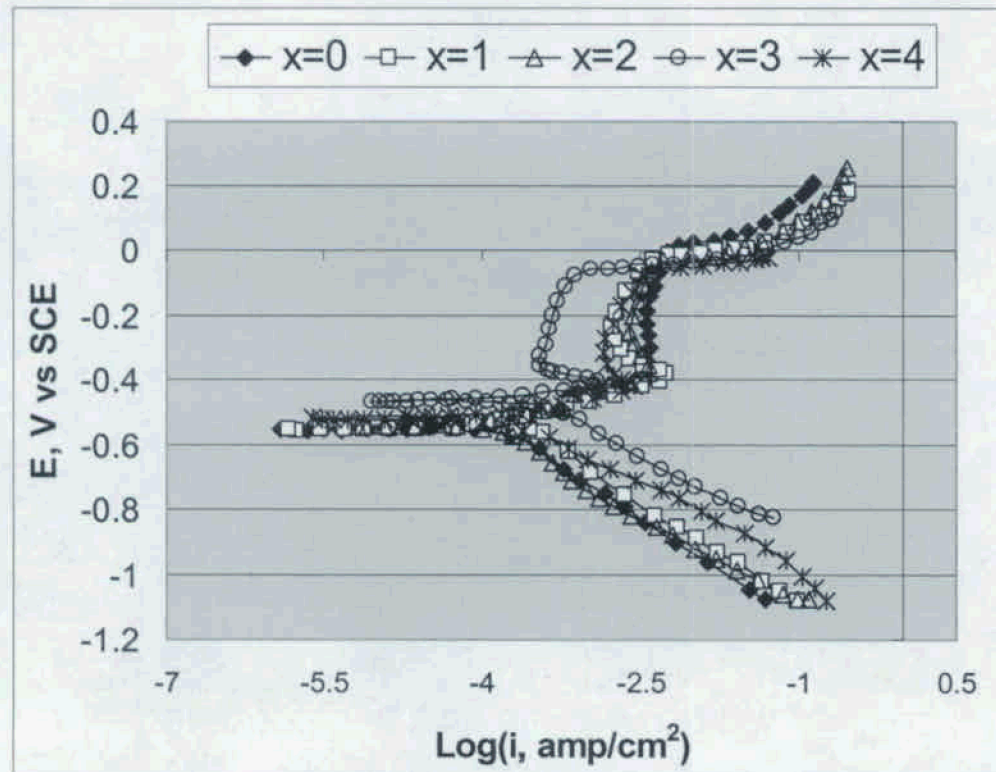
Ref:

[1] B. Seshu, A.K. Bhatnagar, A.Venugopal and V.S.Raja, *Journal of Materials Science*, 32, 2071 (1997).

[2] X. He, J.J. Noël and D.W. Shoesmith, *Corrosion Science*, 47, 1177 (2005).



# Polarization Data on Iron-alloy



Polarization curves for  $\text{Fe}_{68}\text{Ni}_{14-x}\text{Mo}_x\text{Si}_2\text{B}_{16}$  metallic glasses with  $x=0, 1, 2, 3, 4$  in 1N HCl [1].

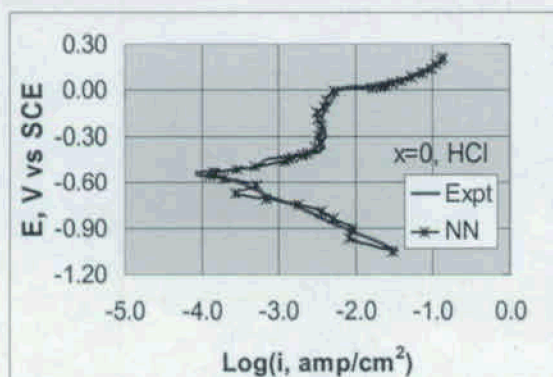
# Polarization Data on Iron-alloy

Kinetic parameters for  $\text{Fe}_{68}\text{Ni}_{14-x}\text{Mo}_x\text{Si}_2\text{B}_{16}$  metallic glasses with  $x=0, 1, 2, 3, 4$  in 1 N HCl [1].  $\beta_a$  and  $\beta_c$  are Tafel constants and  $R_p$  is polarization resistance.

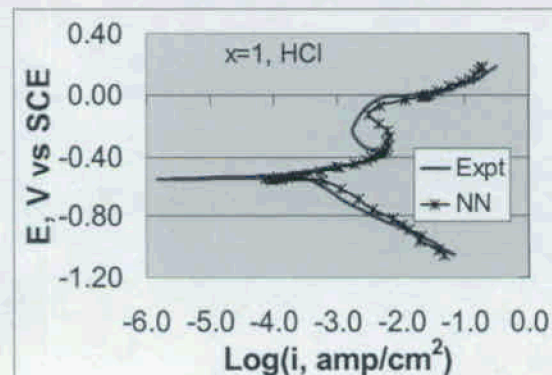
Sample	$E_{\text{corr}}$	$i_{\text{corr}}$	$\beta_a$	$\beta_c$	$R_p$
	mV	$\mu\text{A}/\text{cm}^2$	mV/decade	mV/decade	ohm.cm <sup>2</sup>
1	-571	153	109	203	201
2	-540	226	71	223	103
3	-567	153	79	237	167
4	-506	271	31	154	41
5	-554	141	70	134	142



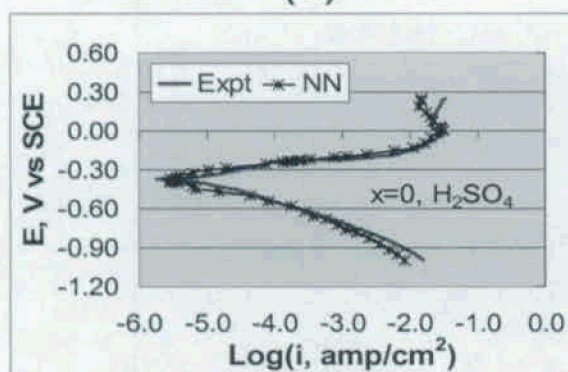
# Analysis of Polarization Data on Iron-alloy



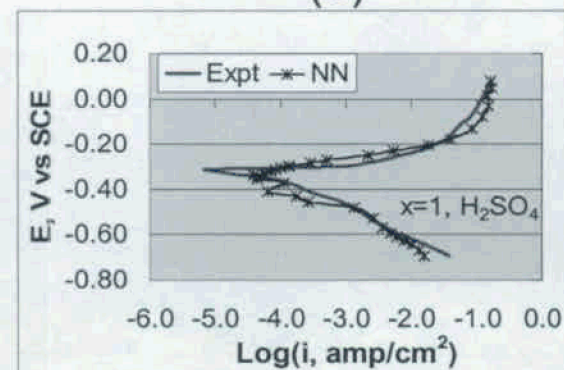
(a)



(b)



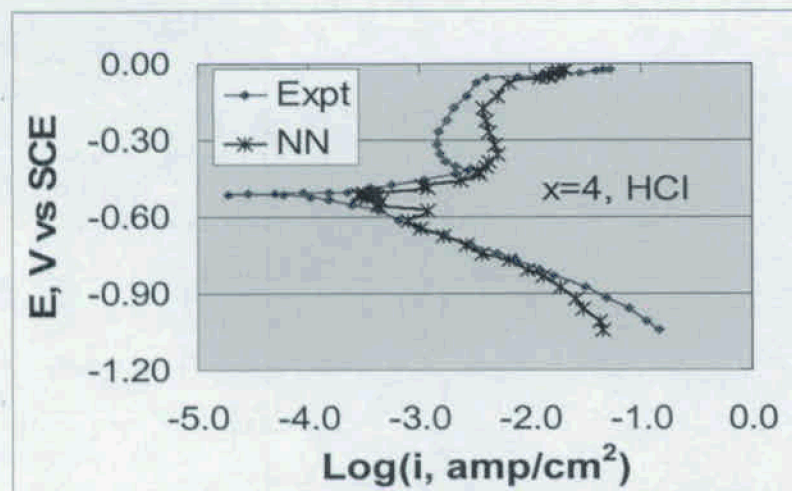
(c)



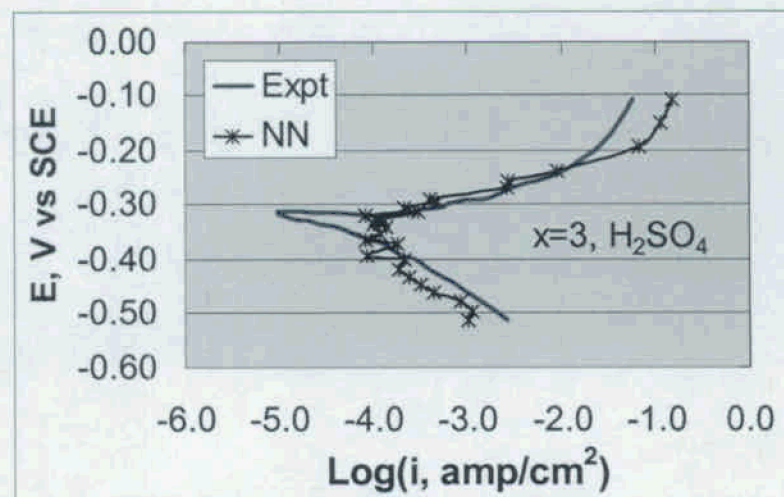
(d)

Polarization curves for  $\text{Fe}_{68}\text{Ni}_{14-x}\text{Mo}_x\text{Si}_2\text{B}_{16}$  metallic glasses with (a)  $x=0$  in 1N HCl, (b)  $x=1$  in 1N HCl, (c)  $x=0$  in 1N  $\text{H}_2\text{SO}_4$ , and (d)  $x=1$  in 1N  $\text{H}_2\text{SO}_4$ . The NN is validated with the same data sets as the model was trained with.

# Analysis of Polarization Data on Iron-alloy



(a)



(b)

Polarization curves for  $\text{Fe}_{68}\text{Ni}_{14-x}\text{Mo}_x\text{Si}_2\text{B}_{16}$  metallic glasses with (a)  $x=4$  in 1N HCl, and (b)  $x=3$  in 1N  $\text{H}_2\text{SO}_4$ . The NN is validated with test data sets that were not seen by the model during training.



# Localized Corrosion Data on Ti-2

Data used to develop a damage function for crevice corrosion of Ti-2 in 0.27 mol/dm<sup>3</sup> NaCl solution at 95°C [2]. Q refers to the charge in coulombs; W is the weight loss expressed in coulombs.

Incubation period (days)	Duration (days)	Maximum penetration depth, Dmax ( $\mu\text{m}$ )	Q (C)	Weight change (g)	Q/W	O <sub>2</sub> consumed (%)
1.19	3.02	140	9.03	0.0525	0.0143	0.5
2.42	9.78	604	106	0.0773	0.114	5
0.54	15.13	938	371	0.1595	0.193	19
0.76	15.5	969	333	0.2728	0.101	17
1.31	23.15	1265	578	0.2333	0.206	29
1.22	31.07	1645	899	0.321	0.232	45
1.26	53.14	1611	1189	0.6508	0.152	60
0.95	64.26	1694	1269	0.6603	0.159	64

# Analysis of Localized Corrosion Data on Ti-2

## Models for localized corrosion analysis:

- Empirical model
  - > Requires a significant database to develop a reliable model
  - > Requires a functional relationship to be specified
  - > Typical model for damage function:  $D = kt^m$
- Deterministic model
  - > Good for future prediction
  - > Decreases the requirement of large databases
  - > Requires that the mechanism of damage functions be known for the specific operating conditions

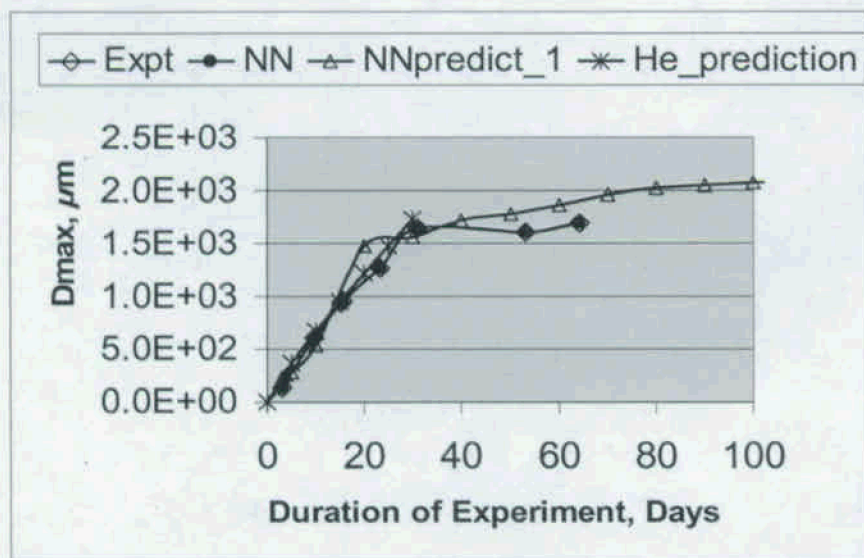


# NN Model for Localized Corrosion Analysis

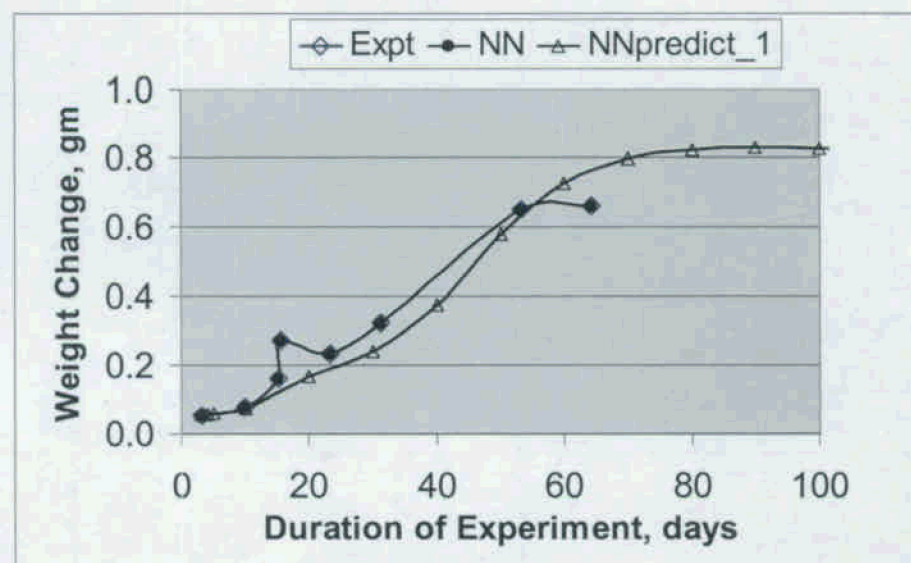
## Supervised NN (backpropagation) model:

- Requires an extensive data base (as empirical model)
- Helps to perform non linear mapping between the inputs and outputs that can be used to establish relationships between input and output parameters/variables
- NN learns the “underlying functions” that relate those mappings and once trained, they can be used to make “smart predictions” or extrapolations over extended parameter range (unlike the ability of empirical models)
- NN backpropagation for the He et al. data
  - 3 layers, 20 neurons/layer, hyperbolic tangent function
  - Inputs: incubation time, duration of experiments
  - Outputs: weight change, Q, Q/W, % O<sub>2</sub> consumed and their logarithms
  - RMS error <0.001%

# Analysis of Localized Corrosion Data on Ti-2



(a)

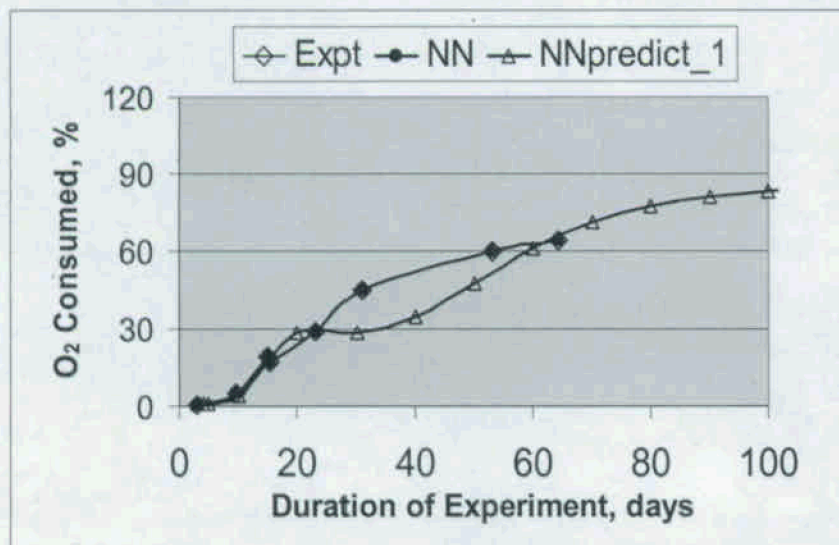


(b)

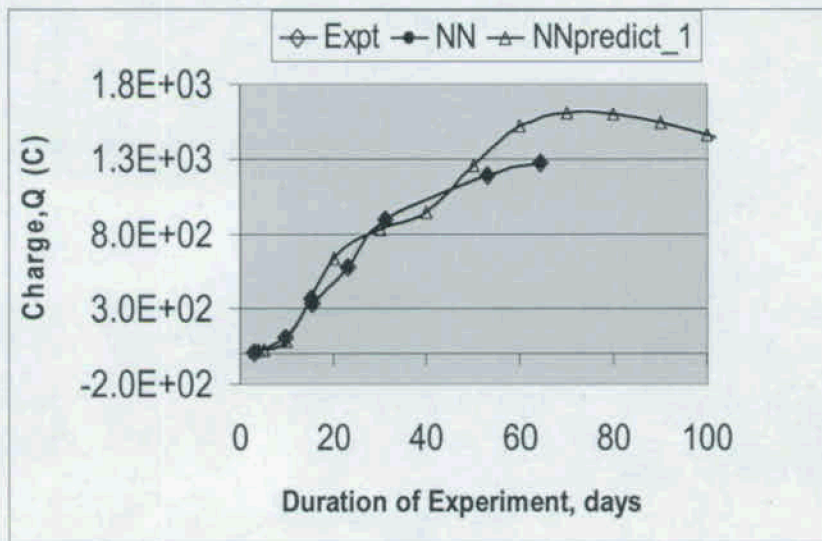
NN results in the study of crevice corrosion of grade-2 Ti in  $0.27\text{mol/dm}^3$  NaCl solution at  $95^\circ\text{C}$  [2]. (a) Prediction of maximum penetration depth,  $D_{max}$ , and (b) Prediction of sample weight change, Expt: experimental data, NN: model prediction using the same input data that was used for training, NNpredict\_1: prediction for 100 days, He\_prediction: predicted  $D_{max}$  by He et al.  $D_{max} = 89.74t^{0.87}$ .



# Analysis of Localized Corrosion Data on Ti-2



(c)



(d)

NN results in the study of crevice corrosion of grade-2 Ti in 0.27mol/dm<sup>3</sup> NaCl solution at 95°C [2]. (c) Prediction of O<sub>2</sub> consumption, and (d) Prediction of electronic charge difference, Q. Expt: experimental data, NN: model prediction using the same input data that was used for training, NNpredict\_1: prediction for 100 days.

# Summary

- Two examples of NN backpropagation are presented
- We demonstrated that this type of technique or tool can represent a useful complement to empirical or deterministic models
- NN model developed using polarization data for general corrosion of metallic glasses shows good agreement with experimental data at the specified operating conditions
- NN model developed using crevice corrosion data on grade-2 Titanium was validated against experimental data to show good agreement. The model was then used to predict future maximum penetration depth and other variables of interest under similar operating conditions